DOCUMENT RESUME

ED 127 344 TM 005 433

AUTHOR Fuller, Edward,
TITLE Self-Instructional Manual for Psychological Assessment. Manpower and Community College Counseling Program.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.; Oregon State Univ., Corvallis. School of Education.; Portland State Univ., Oreg. School of Education.
SPONS AGENCY Bureau of Adult, Vocational, and Technical Education (DHEW/OE), Washington, D.C. Div. of Manpower Development and Training.
PUB DATE Sep 73
NOTE 64p.

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS *Autoinstructional Aids; Correlation; Higher Education; *Manuals; *Measurement; Pretests; Sampling; Scores; Standard Error of Measurement; Statistical Analysis; *Statistics; Test Reliability; Test Validity

ABSTRACT

This self-instructional manual for psychological assessment focuses on the following topics: (1) general statistics, (2) central tendency, (3) random, continuous, and discrete variables, (4) variability, (5) measuring variability, (6) sampling, (7) derived scores, (8) covariation, (9) reliability and validity, and (10) standard error of measurement. A glossary of terms used and pretests of the topics covered are included. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED127344

MCCCP SELF-INSTRUCTIONAL MANUAL for PSYCHOLOGICAL ASSESSMENT

TM005 433

Division of Manpower Development and Training
U.S. Office of Education

2

Northwest Regional Educational Laboratory
Portland, Oregon 97204

OSU The School of Education
Oregon State University
Corvallis, Oregon 97330

PSU School of Education
Portland State University
Portland, Oregon 97207

SELF-INSTRUCTIONAL MANUAL
FOR
PSYCHOLOGICAL ASSESSMENT

September 1973

Manpower and Community College Counseling Program

Edward Fuller
Program Coordinator

Funded by the Manpower Division of the U.S. Office of Education

Northwest Regional Educational Laboratory
700 Lindsay Building/710 S.W. Second Avenue
Portland, Oregon 97204

and

School of Education
Oregon State University
Corvallis, Oregon 97330

School of Education
Portland State University
Portland, Oregon 97207

Northwest Regional Educational Laboratory, 1973

Published by the Northwest Regional Educational Laboratory, a private nonprofit corporation supported in part as a regional educational laboratory by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

CONTENTS

1.0 GENERAL STATISTICS 1
- Descriptive Statistics
- Inferential Statistics

2.0 CENTRAL TENDENCY 6
- Average: An Introduction
- Central Tendency
- Mean
- Median
- Mode
- Advantages and Disadvantages of the Mean, Median and Mode as Representative Values

3.0 RANDOM, CONTINUOUS AND DISCRETE VARIABLES. 10
- Random Variable
- Continuous Variable
- Discrete Variable
- Review

4.0 VARIABILITY. 12
- General Discussion
- Measurement and Variability
- Frequency Distribution
- Distribution Shapes

5.0 MEASURING VARIABILITY. 16
- General Discussion
- Range
- Interquartile Range
- Standard Deviation
- Summary
- Applications
- Review
- Example of Variability

6.0 SAMPLING 20
- General Discussion
- Population
- Sampling
- Random Sampling

7.0 DERIVED SCORES 23
- Percentages
- Percentiles
- Standard Score

8.0 COVARIATION. 25
General Discussion
Correlation
Computation of Rank-Order Correlations
Computation of Product-Moment Correlations

9.0 RELIABILITY-VALIDITY 28
Reliability
Validity

10.0 STANDARD ERROR OF MEASUREMENT. 30

11.0 SUMMARY. 31
General Statistics
Data

12.0 GLOSSARY 32

13.0 PRE-TESTS. 37
Section 1 -- General Statistics 38
Section 2 -- Central Tendency 39
Section 3 -- Random, Continuous and Discrete Variables. . . 42
Section 4 -- Variability. 44
Section 5 -- Measuring Variability. 47
Section 6 -- Sampling 48
Section 7 -- Derived Scores 51
Section 8 -- Covariation. 54
Section 9 -- Reliability-Validity 56
Section 10 - Standard Error of Measurement. 58

1.0 GENERAL STATISTICS

1.1 Descriptive Statistics

Descriptive statistics provides us with many kinds of procedures for summarizing various measured characteristics of a group of individuals or things.

EXAMPLE: Note the column of values labeled X.

Student	X
John	$10
Ralph	5
James	20
Mary	18
Chris	25
Randy	17
Fred	14
Ivan	22
Joan	85

The following can be said about these data by the calculation of simple statistics:

> The total is $216, the arithmetic average or mean is $24, half the scores lie above $18 and half below $18. It is apparent that the number of Joan's dollars is greater than the other amounts.

To guess why Joan has a different amount would involve making an inference and would go beyond a description of the data.

One of the most useful things that descriptive statistics can do for us is to give us a method for analyzing data and portraying the data in a manner that a single number tells how high a particular group stands on a given scale. This summarizing of data always results in the loss of some information and this loss may produce a misleading distortion. For example, the average salary of a company's employees may be $9,000 per year, but the chance of a woman employee being paid more than $5,000 is very unlikely.

EXAMPLE: Assume the two frequency distributions shown in Figures 1.1 and 1.2 represent the scores of two classes on a statistics test.

FIGURE 1.1
Test Scores of Students in a Monday Night Class

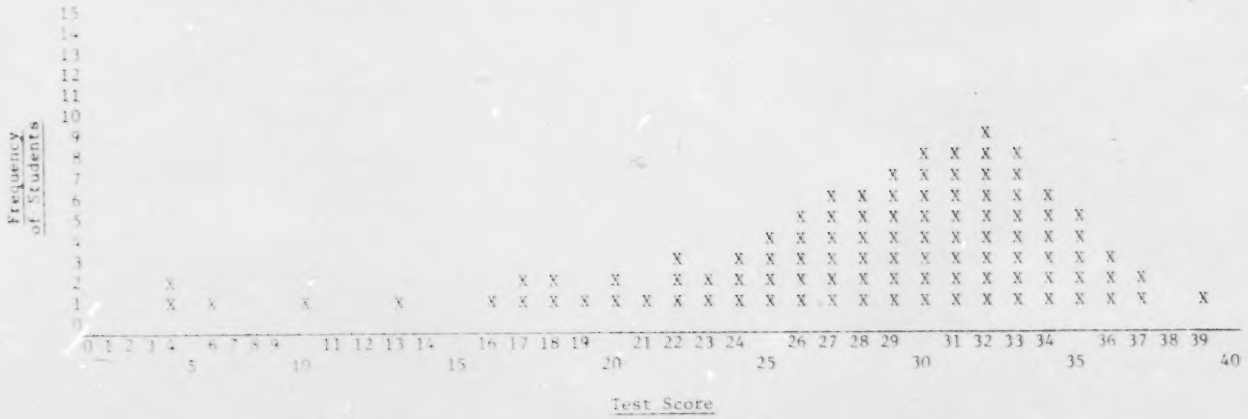
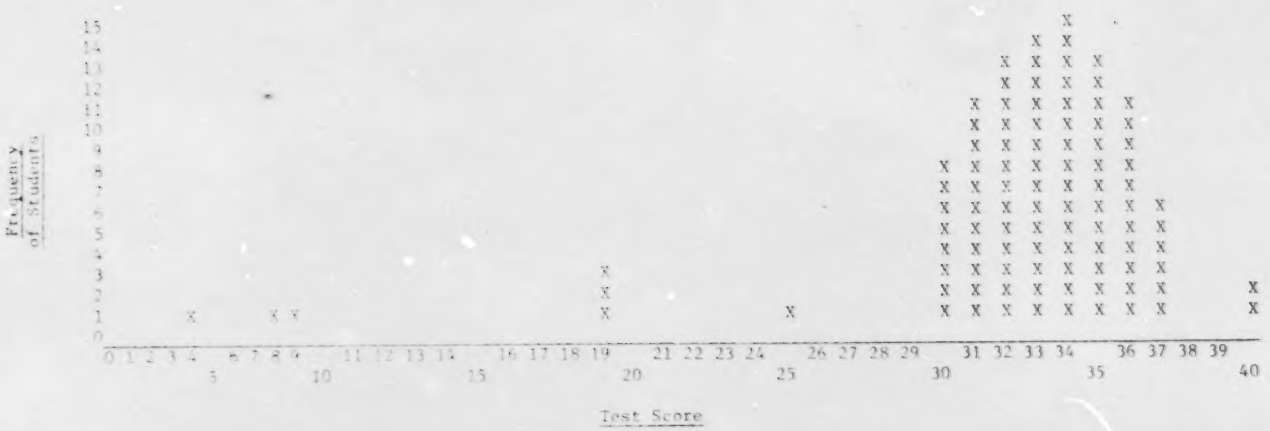


FIGURE 1.2
Test Scores of Trainees in a Wednesday Night Class



What kinds of questions can usefully be asked so that comparisons can be made between the classes?

1. What single score best represents the total pattern of scores for the Monday night class? For the Wednesday night class?

   a. The arithmetic mean of all scores
   b. The median (score above and below which 50% of all the scores fall).
   c. The mode (the most frequent score).

2. In what patterns are the scores distributed, i.e., across what range does any chosen proportion of scores fall?

   a. 100% of the scores
   b. the middle 75% of the scores
   c. the middle 50% of the scores

3. What is the relative position of a score in a distribution of scores? That is, what is the position of a score value relative to the rest of the scores?

   a. for the Monday night class, the 90th percentile score is about _____
   b. for the Monday night class, the 10th percentile score is about _____

Descriptive statistics provides us with methods for describing data in terms of:

1. representative values
2. a pattern in which scores are distributed as shown graphically by a frequency diagram
3. what percentage of scores fall above and below a given score.

These summaries provide information about a given group of scores; they also permit comparisons between different groups.

1.2 Inferential Statistics

Inferential statistics provides us with many kinds of time-saving procedures for finding out about various characteristics of a large or relatively unavailable groups of individuals or things. The saving comes about by calculating descriptive statistics on a sample taken from the larger or relatively unavailable (the population). From what we find out about our sample, we can with greater or less certainty say many things about the individuals or things in the entire group (population) whom we neither observed nor measured.

Inferential statistics can be defined as methods for estimating the true conditions of all data from sample data, and describing the risks of being correct or incorrect in the estimate.

EXAMPLES: Predicting voting behavior for a forthcoming presidential election based on a small random sample of all registered voters. Predicting criterion performance of applicants to a training program on the basis of their aptitude test scores. The prediction formula is based on the relationship between the aptitude scores and criterion performance of previous applicants of similar age and experience.

Inferential statistics is based on the laws of chance. Many introductory statistic books include a section or chapter on probability. However, students usually fail to develop an understanding of inferential statistics by learning such facts as the following: If a card is randomly selected from a deck, the chances are 13 in 52 or 1/4 that it is a spade. If one does not play poker or shoot dice, examples used to explain statistics which use cards or dice often fail to clarify the point being made. Hopefully, the current form of reporting the weather forecast will make individuals more aware of the likelihood or degree of certainty of chance events. If the weather forecaster says there is 100% chance of rain, it is highly likely that measurable rain will fall. If he says 10% chance of rain, many people will leave their umbrellas at home, and on the average, they will have guessed right nine times out of ten.

Inferential statistics consists of procedures for making decisions about hypotheses based on the analysis of sample data.

EXAMPLE: Mrs. Jones often gets a sinus headache when the weather changes from clear and dry to cloudy and rain. For the past six months she has kept track of whether it rains following the onset of her headache. Headache has been followed by rain 22 times of the 70 times it has rained. Mrs. Jones had 12 headaches which were not followed by rain. This information is presented in the following table:

	No Rain	Rain	Total
Headache	12 times	22 times	34 times
No Headache	76 times	48 times	124 times
Total	88 times	70 times	158 times

Hypothesis: There is no relationship between Mrs. Jones' sinus headaches and likelihood of rain.

Decision: Should her husband take an umbrella on the mornings his wife wakes up with a sinus headache?

2.0 CENTRAL TENDENCY

2.1 Average: An Introduction

Imagine a column of data such as test scores obtained by giving the Wonderlic Personnel Test to all job applicants registering during the month of July. How would you go about determining the sirgle score which best represents all the scores in the column. Until some better method is discovered, you would most likely compute some form of average. This might be the mean, or the median, or possibly the mode. An average value is intended to be representative of the level of all scores. An average expresses the outcome of the plus and minus effects of many influences on what is observed. Some of these influences are constant from one observation to the next. The constant influences together with the cancelling out tendencies tend to make scores cluster around an average score. This clustering is called central tendency.

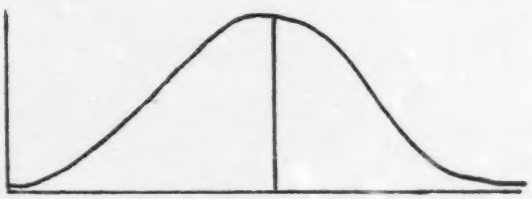
Thus, two questions can be asked of the above mentioned test scores:
1. What is the average test score?
2. To what extent do the other test scores differ from this average?

2.2 Central Tendency

Definition: Central tendency is the propensity of scores to cluster around an average.

Any score or measurement value is usually the result of many independent influences. Some influences act to increase the score, others act to decrease the score. Thus, there tends to be a balancing out or equalizing effect that stabilizes the variable being measured around some central value.

EXAMPLE: The classical example of central tendency is the clustering of scores around the mean of the bell-shaped, normal curve.



The bell-shaped distribution of scores is more likely to be found with a large number of observations (1,000 or more) and less likely to be found with a small number of observations (10-30).

Summary: Randomly selected scores tend to bunch around some central value. The causes of such distributions (like the bell-shaped curve) result from independent factors which, occurring randomly, tend to balance each other out around some stable level of the variable being measured.

2.3 Mean (a term that is commonly referred to as the average, but in reality the average value can be either the mean, median, or mode)

Definition: The arithmetic mean is the sum of a set of scores divided by the number of scores.

Equation: In equation form the mean can be expressed as: $\bar{X} = \frac{\Sigma X}{N}$

Where $\bar{X}$ is the arithmetic mean of X scores:

Σ means "the sum of"

X is each raw score

N is the total number of scores

The equation can be read as: "X-bar" (the mean) is equal to the sum of X (adding each raw score) divided by N (the total number of scores).

EXAMPLE: Table 1 lists the scores of a group of job applicants on an arithmetic test. The mean of 27 is a measure of the central tendency in these data. Table 2 lists their monthly salary on other jobs for the past month. In this case the mean of $383.12 is misleading. No applicant earned this amount. Furthermore, six of the eight earned less than this amount.

	Table 1		Table 2	
	15		$ 150	
	10		240	
	20		325	
	25		375	
	35		775	
	41		1200	
	45		0	(unemployed)
	25		0	(unemployed)
Total	216	Total	$3065	
Mean	27	Mean	$ 383.12	

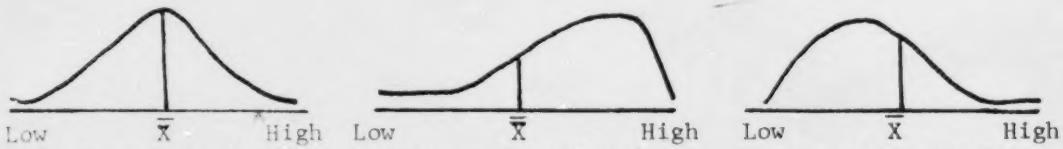
This second example points out two important characteristics of statistics:

1. Following the rules for calculating statistics does not insure useful results.

2. The usefulness of statistics depends upon purposes, decisions, and actions outside of the data and the procedures used in calculating those statistics.

Summary: The arithmetic mean may provide a useful representative value if the scores do indeed cluster around the mean value; are relatively balanced on both sides of the mean; and do not have exceptionally high or low values relative to the mean.

Review: The arithmetic mean is the sum of scores divided by the number of scores. It is one of three kinds of representative values. It is valuable where scores are fairly symmetrical around the mean, but is misleading if the distribution is skewed.



2.4 Median

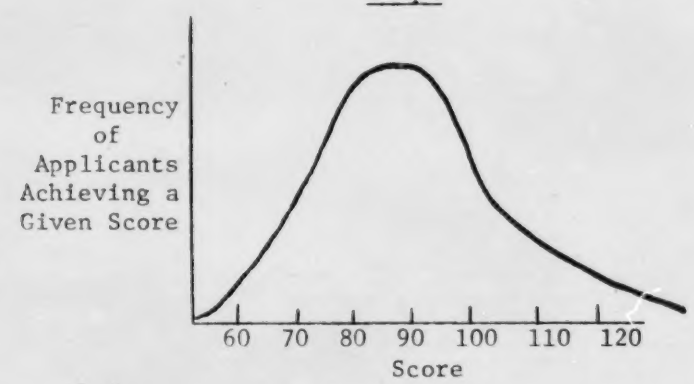
Definition: The median is the middle score--the one that above and below which one-half of the scores fall. The median divides the scores into two halves, each containing 50 percent of the scores.

EXAMPLE: The median may be the preferred method of describing the representative value in a skewed distribution, as shown below:

Number of Applicants	Score
5	60-69
20	70-79
30	80-89
40	90-99
20	100-109
10	110-119
5	120-

Table

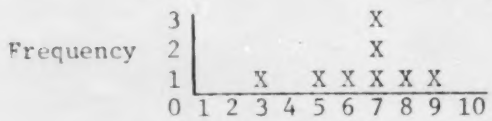
Graph



2.5 Mode

Definition: The mode is the score in the distribution that occurs most frequently.

EXAMPLE: The mode for the following distribution of eight scores is 7. 3, 5, 6, 7, 7, 7, 8, 9



The mode may be preferred in situations where one is interested in the most frequently occurring score, or where one wants a measure of central tendency which can be determined with the least amount of calcuation.

2.6 Advantages and Disadvantages of the Mean, Median and Mode as Representative Values

1. Only the mean can be weighted and averaged with other means and manipulated algebraically.
2. The mean may be misleading if the distribution is highly skewed because of some very extreme cases of one end of the distribution.
3. The median is unaffected by extreme cases, but it cannot be manipulated algebraically.
4. The mode is the most probable and in that sense the most typical. It tends to fluctuate more from one sample to another than the median or mean.

3.0 RANDOM, CONTINUOUS AND DISCRETE VARIABLES

3.1 Random Variable

Definition: A random variable is one whose numerical value is not controlled by the experimenter when he draws samples from a population.

EXAMPLES:
1. Heads or tails from a tossed coin.
2. Numbers in a table of random numbers. The license numbers of cars passing a particular intersection.

A variable ceases to be random when the experimenter exercises control over the appearance or magnitude of the variable.

3.2 Continuous Variable

Definition: A continuous variable is one with a continuous gradation of magnitude.

EXAMPLES:
1. Measurement of a scale of time, distance, area, volume, weight, temperature.
2. Intelligence, aptitude, ability, skill, achievement.

3.3 Discrete Variable

Definition: A discrete variable is one in which changes occur in all or none steps.

EXAMPLES:
1. Heads or tails of coins, numerical values provided by dice and the roulette wheel, number of light bulbs in inventory.
2. Winners, losers; people present or absent; hired or fired.
3. Male or female, alive or not alive.

3.4 Review

A random variable is one whose numerical value is not controlled by the experimenter when he draws samples for observation (e.g., selecting subjects for psychological testing, by randomly choosing names from a campus telephone book listing student names). A continuous variable has a continuous gradation of magnitude (blood pressure, heights of first graders, weights of Army recruits, measurements of Miss America contestants). A discrete variable changes in all-or-none steps from one category to another (sex of high school graduates, national origin of emigrants, rights vs. wrongs on a single achievement test item.

It is important to remember that statistical procedures appropriate for continuous variable data usually should not be used on discrete variable data. Exceptions to this general rule can be made when the discrete values are assumed to be a measure of an underlying continous variable. Number of items correct on a test is an example of a

discrete variable, but presumably the test measures performance which is a continuous variable. Statistical procedures appropriate for discrete data can be used on continuous data if these data are artificially categorized into such categories as pass-fail; above average, average, below average; age groups such as 20-29, 30-39, 40-49, 50-59, etc.

4.0 VARIABILITY

4.1 General Discussion

Without variation in a set of measurements or objects, statistics would be meaningless. Differences between measurements and in the conditions that produce or are related to these differences are the subject matter of statistics.

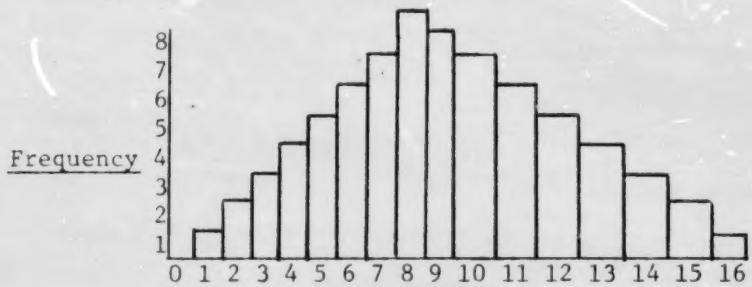
There are a variety of scales for measuring variability. One is the two-category scale (yes/no, present/not present) which measures a discrete variable. Other scales measure continuous variables such as time, height, distance, etc.

In using statistics one usually wants to separate those changes that are constant, regular and predictable from those that are random.

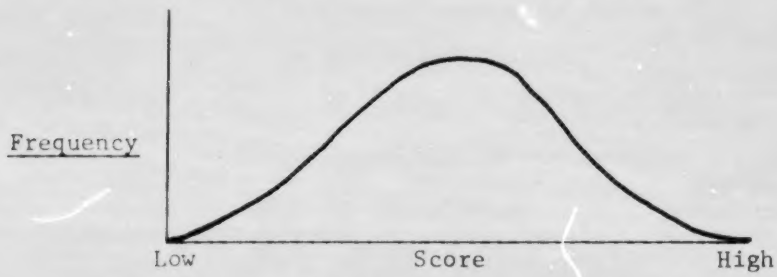
4.2 Measurement and Variability

Definition: Variability is the amount of scatter or dispersion among a set of scores obtained from the use of a given measurement procedure.

EXAMPLE: Sixteen coins tossed a thousand times, and the number of heads observed is recorded and portrayed graphically below:



The performance of a student on 30 standardized intelligence tests is likewise recorded and displayed graphically:

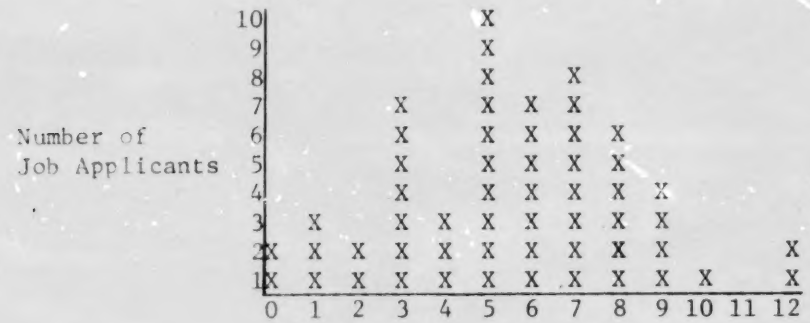


Variability is a measure of how much conditions change from one determination to the next. The amount of variability is related to the number of independent and random influences that occur in the variable being measured. The more complex the variable, the more variable it is likely to be. Scores can vary even though the variable being measured is constant because the measurement procedure being used is not consistent.

4.3 Frequency Distribution

Definition: A frequency distribution is a plot of scores (values) against the number of times each score (value) appears.

EXAMPLE: The figure below shows the number of errors made by a group of job applicants on a motor skills test. Each "X" represents a job applicant; thus, there were two who made zero errors, three who made one error, etc.



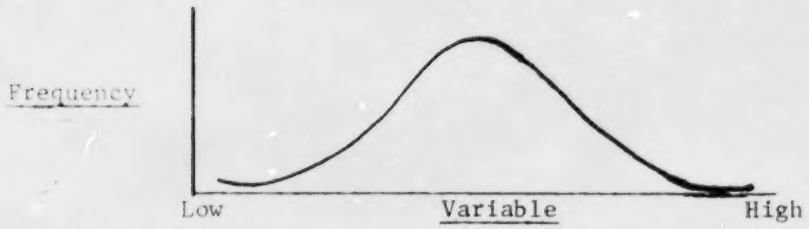
A frequency distribution in the form of a graph is a simple yet impressive tool in analyzing and presenting any set of data. Useful hypotheses can often be generated by observing graphic displays of distributions.

4.4 Distribution Shapes

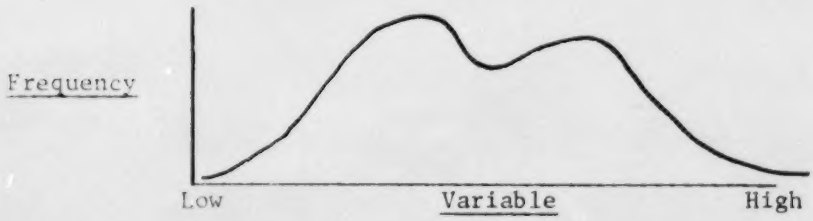
The average or representative value of a group of scores (the mean, median, or mode) rarely conveys information about the processes generating the numbers. The manner in which the numbers are distributed around the average can allow interpretations to be made from graphical displays of these distributions.

EXAMPLE: Some suggestions of the meanings of distribution shapes are listed below:

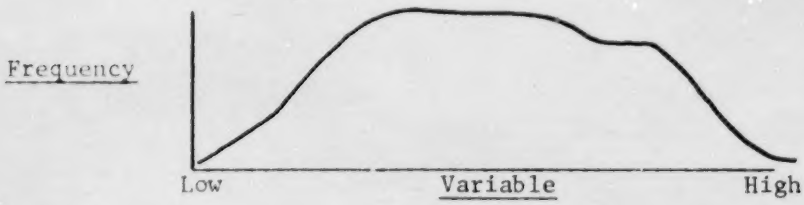
1. The normal frequency curve suggests that many independent influences affect the variable and that the differences in values were determined by random influences.



2. A bimodal distribution curve suggests two populations were sampled, each with a different average amount of the variable being measured.

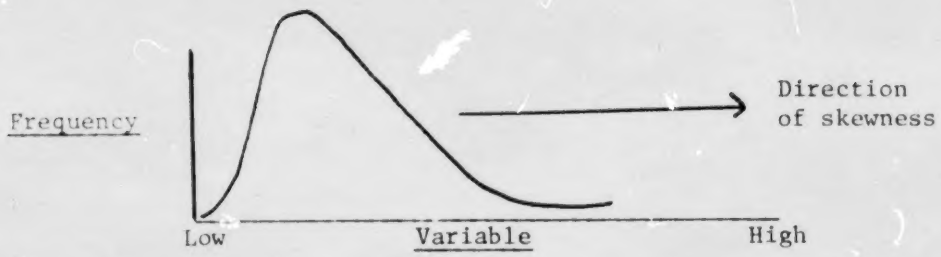


3. A flat distribution curve indicates that the sample lacks a well defined central tendency. If time was the variable, the process being measured has brief starting and stopping characteristics and a long steady period. Multiple choice and true-false tests often produce a raw score distribution which approaches a rectangular shape.



4. The following skewed distribution curve suggests one or more of the following:

   A. A scale with insufficient range at the lower end.
   B. A selective influence on data at the left end of the scale.
   C. Some arbitrary cut-off point on the variable at the left end.
   D. It is a time measurement, which often has a long tail to the right.



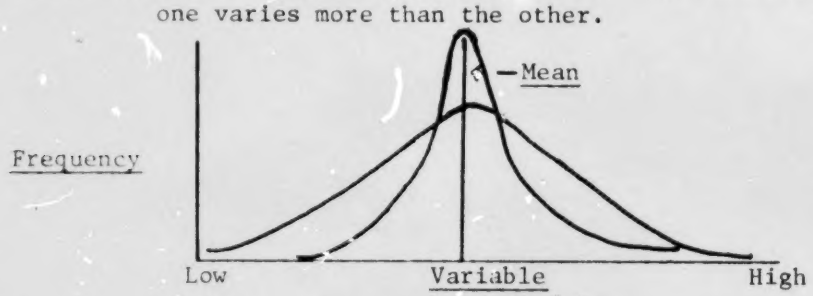
Review: All data do not fit the normal frequency curve, curves may be skewed, bimodal (with one or more high points), or relatively flat.

5.0 MEASURING VARIABILITY

5.1 General Discussion

It is usually important to know how a group of measurements are distributed around the average value. A method of measuring the dispersion of measurements is needed, so that one can determine if two distributions with the same mean differ in the manner in which measures fluctuate around that mean.

EXAMPLE: The two distributions below have the same mean, but one varies more than the other.



5.2 Range

The total range across which a group of measurements is spread is one indication of variability. The range is unreliable, however, in that it can be affected by unusual measures at the extremes of the distribution.

5.3 Interquartile Range

To reduce the chances of being misled by unusual occurrences, a range can be established through the middle 50 percent of scores or the middle 75 percent of scores, or some other arbitrary range.

5.4 Standard Deviation

If the distribution meets certain requirements, a statistic called the standard deviation can provide a measure of variability that is more consistent than the range. The standard deviation employs the mean as a reference point and measures the amount of dispersion from that point. It is a valuable tool in that it permits comparisons to be made of the variability of one group of measures with another.

5.5 Summary

A group of measurements can often be adequately described in terms of:

1. Some form of average or representative value (mean, median or mode) plus
2. Some measure of the variability of the measures (the dispersion of the scores around this average value).

Measures of central tendency and variability are fundamental requirements in:

1. Evaluating scores within a distribution.
2. Assessing differences between distributions.
3. Making predictions about unknowns from known data.

5.6 Applications

How can the distributions of scores for test A and test B be compared?

1. Range: By measuring the distance between the highest and lowest scores in the distribution. But a single extreme score can mislead the observer with just this information. In addition, there is no information about where the scores cluster.

2. Interquartile Range: By dividing the scores into a lower 25 percent (from the first quartile to the second), the next 25 percent (from the second quartile to the third) and so on, you can obtain the interquartile range which is the difference in score between the score for the 25th percentile and the score for the 75th percentile.

3. Average Deviation: By measuring the deviations from the average value (such as the mean), by subtracting every score from the average value, disregarding the negative signs (values) and dividing by the total number of scores. But the average deviation does not enable one to compare different parts of a distribution with itself or other distributions, and is thus, rarely employed.

4. Standard Deviation: When the distribution is fairly symmetrical around its mean, and does not have arbitrary cut-offs, it is possible to calculate a measure of variability that "best-fits" the characteristic of the frequency curve, the statistic being a standard deviation.

5.7 Review

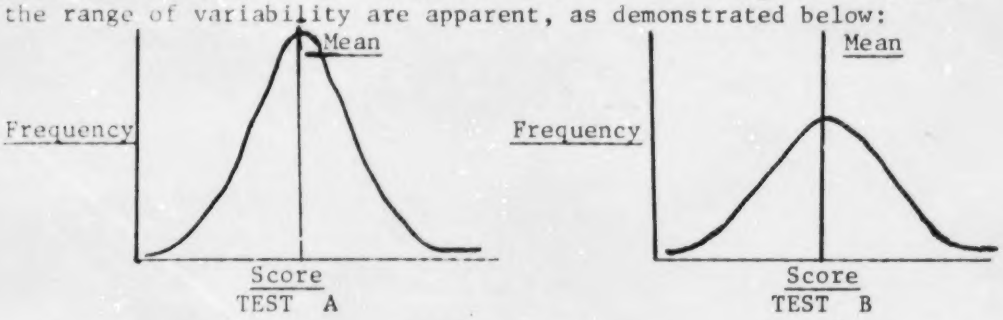
Range: Is the range on the variable from the highest to lowest score.

Quartile Ranges: Gives the interval on the variable on which successive 1/4's (25%'s) of score frequencies fall from highest to lowest. The interquartile range is the range of scores of the middle 50% of the cases.

Standard Deviation: Provides a measure of dispersion that enables comparisons between sets of data as to amount of variability. It is usually the most useful measure of variability.

5.8 Example of Variability

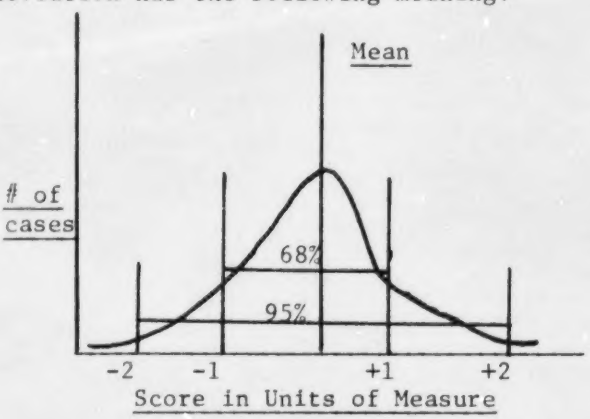
A group of students are tested for clerical aptitude on two tests, A and B. The mean score for test A is identical to the mean score on test B. But looking at the frequency distribution, large differences in the range of variability are apparent, as demonstrated below:



A formula exists for calculating the dispersion of a group of measures about their mean. It is called the standard deviation formula. When the standard deviation formula is applied to data that are normally distributed (bell-shaped distribution curve) about their mean, the resulting value of the standard deviation has the following meaning:

A. Sixty-eight percent of the scores fall within + or - 1 standard deviation of the mean.

B. Ninety-five percent of the scores fall within + or - 2 standard deviations of the mean



In the example of the clerical aptitude test mentioned above: If the standard deviation of test A is 10 (units of measure), it means that 68% of the students (cases) fall within + or -1 standard deviation of the mean score for that test.

If the standard deviation of test B is 20 (units of measure), it means that 95% of the students (cases) fall within + or - 40 points of the mean score for that test.

6.0 SAMPLING

6.1 General Discussion

Usually it is necessary to draw conclusions about a large number of cases from relatively few cases.

EXAMPLE: To determine the average blood pressure of all adult males in the U.S.A., it is more efficient to measure a few thousand U.S. males and estimate or infer the average blood pressure of all U.S. males from the smaller group average. The question then becomes, how are the few cases selected so that legitimate conclusions about the larger group can be made? The answer to this question is to use sampling procedures.

First, it is necessary to determine what the population is, in the case above, "all U.S. males." Once the population is identified, samples are drawn from it that will realistically reflect the attributes of the population that are to be measured. To avoid error, attempts are made to minimize bias or selectivity that would give preference to some portions of the population over other portions. One way to avoid bias is to select cases randomly. Cases are selected randomly when each case in the population has an equal opportunity to be selected for the sample.

Even in unbiased samples, chance can produce results that are not representative. For example, when a normal coin is tossed, it can come up five heads in a row. It is possible, however, to account for the probability of unusual events affecting the sample, and indicate these as the risk one takes as to whether the random sample is biased. To reduce the risk (that is, to be more certain that the conclusion is a true one), larger samples can be drawn. Sampling the entire population reduces this risk to zero; however, this procedure is often costly and unnecessary, as relatively small samples can provide conclusions that carrry little risk. The variability of the attribute in the population being mesured determines the size of the sample necessary to make low risk conclusions. For example, since the weight of adults is more variable than their height, more samples of weight are needed to be sure their average weight than would be needed to be sure of their average height.

6.2 Population

Definition: A population is all of the individual objects and events from which sample data have been drawn. When drawing samples with the intent to generalize to a population, careful definition of the population is required.

EXAMPLE: An adult male (an individual object), having his blood pressure taken (an event), his actual blood pressure (a characteristic or attribute).

There are several ways of defining and measuring the blood pressure of individuals. One way is to use a sphygmomanometer and take his blood pressure after he has rested quietly for 15 minutes. The usual way is to measure his blood pressure without the preliminary rest period. The results will be different. It is necessary, therefore, to define the population in a way that conditions of measurement are known. Also, it is important to know to whom the findings are to be generalized. If sampling to generalize about the blood pressure of all adult U.S. males, this would be different from all U.S. males.

Complete definitions of populations should therefore characterize the methods of observation or measurement and the group to which some conclusion is intended to apply.

6.3 Sampling

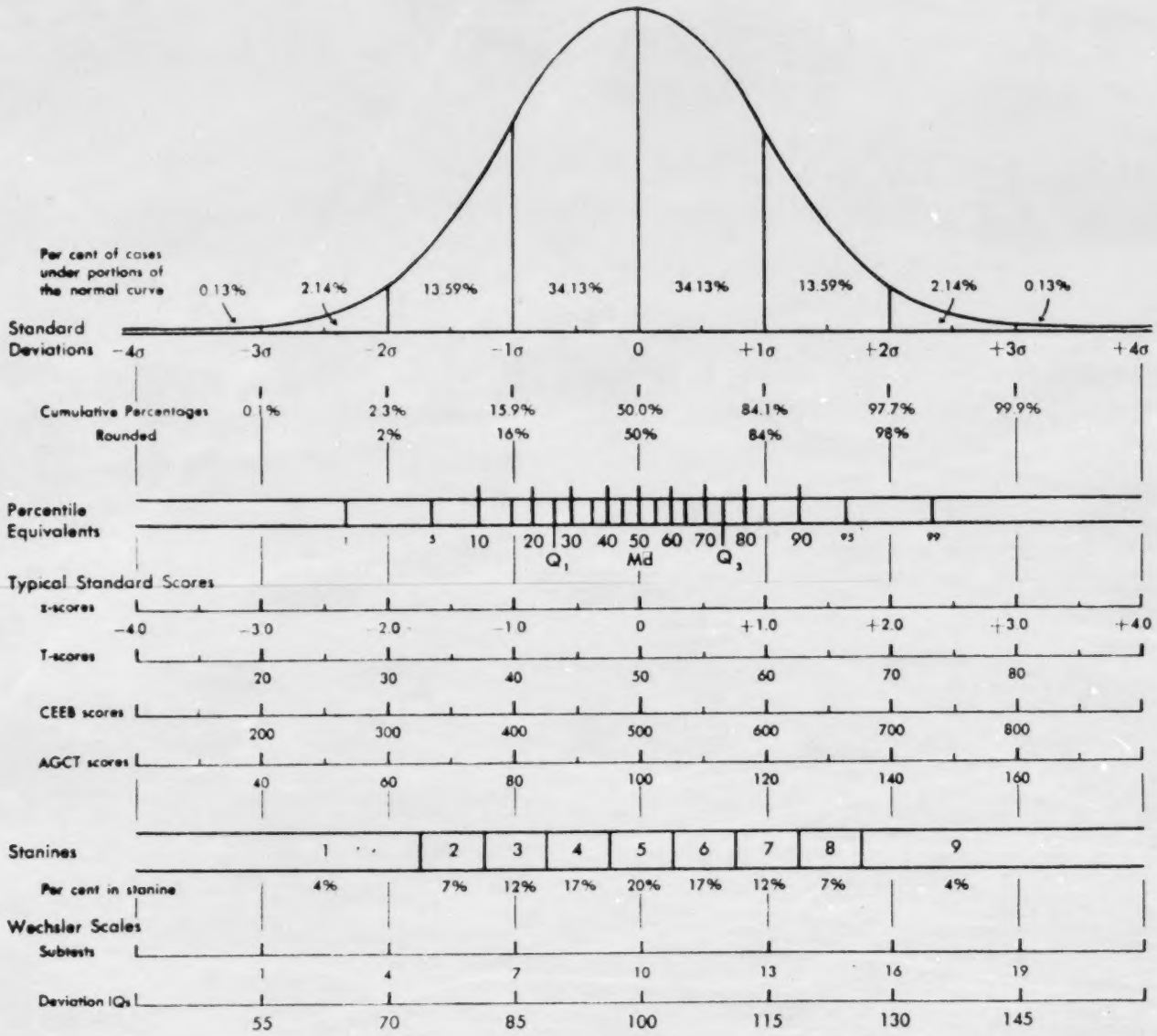
Definition: A sample is:

1. One or more observations or measurements
2. of selected individuals, objects, or events
3. made with the intent to generalize
4. to some intended population

The purpose of sampling is to estimate one or more characteristics or attributes of a population from some of its members.

6.4 Random Sampling

If each member in a group of objects has the same chance of being drawn for the sample as any other member, the sampling of objects is random. Random sampling can be done by identifying every individual in the population, assigning each individual a numbered token, thoroughly mixing the tokens in a box and drawing the tokens out randomly. Practically, many compromises are made, some of which can produce bias; therefore, it is necessary to be aware of the risks taken. A biased sample is one in which systematic deviations are made between the sample objects and the population objects. For example, estimating the blood pressure of U.S. males by using hospital patients as the sample.


Per cent of cases under portions of the normal curve
0.13%
2.14%
13.59%
34.13%
34.13%
13.59%
2.14%
0.13%
Standard Deviations
−4σ
−3σ
−2σ
−1σ
0
+1σ
+2σ
+3σ
+4σ
Cumulative Percentages
0.1%
2.3%
15.9%
50.0%
84.1%
97.7%
99.9%
Rounded
2%
16%
50%
84%
98%
Percentile Equivalents
1
5
10
20
30
40
50
60
70
80
90
95
99
Q1
Md
Q3
Typical Standard Scores
z-scores
−4.0
−3.0
−2.0
−1.0
0
+1.0
+2.0
+3.0
+4.0
T-scores
20
30
40
50
60
70
80
CEEB scores
200
300
400
500
600
700
800
AGCT scores
40
60
80
100
120
140
160
Stanines
1
2
3
4
5
6
7
8
9
Per cent in stanine
4%
7%
12%
17%
20%
17%
12%
7%
4%
Wechsler Scales
Subtests
1
4
7
10
13
16
19
Deviation IQs
55
70
85
100
115
130
145

7.0 DERIVED SCORES

7.1 Percentages

Definition: A percentage is a decimal expression of the proportion of a part to the whole multiplied by 100. To say that 27 of 540 adults are over six feet tall is the same as saying 5% of the adults are over six feet tall (27/540 = 1/20 = .05% = 5%). A percentage is a proportion of some subgroup to the total group. The percentage is thus helpful or misleading according to how clearly the total group (the denominator) is defined.

EXAMPLE: Experimental condition "A" produced a one percent error rate. Experimental condition "B" produced a 1 1/2 percent error rate. Conclusion: The error rate for condition "B" was 50% higher than condition "A". There was a difference of 1/2% in the error rate for conditions "A" and "B". The error rate was only 1/2% higher in "B" than in condition "A".

These differing and misleading conclusions are accurate but incomplete-without specifying the reference base on which the percentage is calculated.

Summary: A percentage is the proportion of a part to a whole converted to a range of values from 0 to 100. One-hundred percent includes the total group. It is useful for comparing fractional values having different denominators.

7.2 Percentiles

The curve on the preceeding page is called the normal curve. Observe that the curve is symmetrical, that there are few scores at the extremes and most pile up in the center. Also, the mean, median and mode are identical. In a normal curve over 99% of the scores fall between $\pm$ 3 standard deviations from the mean; 95% $\pm$ 2 standard deviations; 68% $\pm$ 1 standard deviation.

The percentage of individuals that equal or exceed a given score is called a percentile. If Jim does as well as or better than 50% of his class he is at the 50th percentile of his class, he is also at the median of the class as the 50th percentile by definition divides the class into two equal halves. To find the 75th percentile it would be necessary to count up 75% of "N" or .75 times (x) "N." Where distributions have been divided into tenths, deciles are formed. The lowest tenth of the cases is called the first decile. Quartiles divide into fourths, the first quartile containing the lowest 25% of the scores.

Percentiles represent ordinal measurements because the differences between percentile points are not equal. For example, between the 5th percentile and the 10th percentile there may be only a very few scores (see the normal curve), whereas between the 60th and 70th percentile a great many scores appear.

7.3 Standard Scores

Definition: Standard scores are scores which have been expressed in standard deviation units.

EXAMPLE: Instead of saying that John has an IQ of 115 on the WISC, it is possible to report that John is one standard deviation above the mean. In the same manner, Ralph is also one standard deviation above the mean because he scored 600 on his college entrance examination, where the mean is 500 and the standard deviation is 100.

z-score:

Definition--A z-score indicates how far from the mean an individual has scored in standard deviation units.

EXAMPLE: A z-score of +1.0 means that the subject scored one standard deviation above the mean, a z-score of -2.5 indicates a score 2 1/2 standard deviation below the mean. Scores below the mean yield negative z-scores, scores above the mean yield positive z-scores.

Formula--The formula for a z-score is:

$$z = \frac{X - \overline{X}}{\text{standard deviation}}$$

where: X = raw score

$\overline{X}$ = mean of raw scores

On an examination Fred received a score of 30, the mean score for the exam was 35 with a standard deviation of 10, his z-score, therefore, was:

$\frac{30 - 35}{10}$ or $-.5$ He is 1/2 of a standard deviation below the mean.

T-score: To eliminate negatives and decimals the T-score involves multiplying each z-score by 10 and adding 50.

EXAMPLE: An individual having a z-score of -.5 would have a T-score of $(-.5)(10) + 50 = 45$.

8.0 COVARIATION

8.1 General Discussion

It is often useful to be aware of associations between variables. For example, the question can be asked, What is the relation between the I.Q. of an individual and his average annual income?

If changes in the amount of one variable determine changes in the amount of the other variable, a perfect correlation exists; but when the variables are unrelated to each other, the correlation is zero. The value of a correlation between variables is the extent to which the amount of one predicts the amount of the other.

Low or zero correlations often occur between variables that by common sense would appear to be strongly related to each other, but in fact, have no relationship.

Users of correlational data must remember that a correlation only describes a relationship between variables; it does not prove a causal relationship between variables, although a zero correlation may indicate the absence of a causal relationship.

8.2 Correlation

Definition: Correlation is the extent to which a value of Y variable can be predicted from an X value from distributions of paired X and Y scores.

Perfect Correlation: Five high school students ranked according to test results on the Scholastic Aptitude Test (SAT) were subsequently ranked according to grade point average (GPA). Each student had the same ranking on the SAT as in terms of GPA. This is perfect correlation. Each students' rank in terms of SAT would predict his GPA rank and vice versa.

Zero Correlation: Assume that luck alone determines GPA. This means that GPA rankings will be random and hence unpredictable. This would be zero correlation. Knowledge of a students' SAT score would not provide a basis for predicting his GPA.

Intermediate Correlations: Correlations range from -1 to +1. Both +1 and -1 correlations are perfect. In intermediate correlation the two variables are related, but one set of values can be used to predict only approximately the other set of values.

Continuous-Variable Data: If the data were not rankings but scores on a test, they would be continuous data. Most correlations deal with continuous variable data, such as time, distance, and similar equal-unit scales.

Discrete-Variable Data: Measures of association can be calculated for discrete-variable data, such as, the association between sex of applicant and passing or failing a training course.

EXAMPLES: (of questions involving correlation)

1. What is the relation between the number of questions a student asks per class hour and the instructor's rating of each student?
2. Can the extent to which the number of jobs available in a certain field be predicted by the number of applicants for such employment?
3. Are the amount of hours spent studying positively related to success in school?
4. What is the relation between scores on a civil service test and subsequent performance on the job?

Review: Correlation in paired sets of measures is the extent to which X values predict Y values and conversely, the extent that Y predicts X. Correlation coefficients range from -1.00 to 0 to +1.00.

8.3 Computation of Rank-Order Correlations

In determining whether two set of measures are related, it is sometimes easier to rank individuals from highest (rank = 1) to lowest and then work with these ranks. Where the data are ordinal in nature or if N is less than 25 it is more proper to use rank-order correlations.

Using the earlier example of GPA and SAT scores the following table can be generated:

Student	GPA	SAT	GPA Rank	SAT Rank	D (Difference)	D^2
Ralph	3.5	400	1.5	5	3.5	12.25
Tom	3.5	500	1.5	2.5	1.0	1.00
Dennis	3.0	500	3	2.5	.5	.25
Mary	2.5	475	4	4	0.0	0.00
Steve	2.0	700	5	1	4.0	
					$\Sigma D^2 =$	16.00 / 29.50

Both Ralph and Tom had 3.5 GPA's and should have received a rank of 1 but this would have given Dennis a rank of 2. By giving Ralph and Tom the average rank of 1 and 2 or 1.5 the problem is resolved. The same procedure is applied to SAT scores for Tom and Dennis, where each received the average rank of 2 and 3 ($\frac{2+3}{2}$) or 2.5. The differences between the ranks are determined and placed in column D. These differences are squared and then summed to yield $\sum D^2$. The formula for rank-order correlation is:

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2-1)}$$

Where: ρ = rho = rank-order correlation

$\sum D^2$ = sum of the differences squared

N = the number of pairs of observations

Thus, for the example:

$$= 1 - \frac{6 \times 29.50}{5(5^2 - 1)} = 1 - \frac{177}{120} = 1 - 1.48 = -.48$$

8.4 Computation of Product-Moment Correlations

Product-moment correlations provide a more accurate picture of the relationship than does the rank-order method, as the actual numerical value is employed. To employ the product-moment correlation, it is necessary to have data in interval or ratio scales of measurement with at least 25 pairs of scores.

9.0 RELIABILITY-VALIDITY

9.1 Reliability

Definition: Reliability is the consistency or stability of a measurement.

EXAMPLE: An estimation of reliability can be obtained by computing the correlation between scores from two administrations of the same test to the same persons.

Test scores are considered perfectly reliable when each subject maintains his same rank in a group which takes the same test twice. The closer the value of the correlation approaches 1.00 the more reliable is the test and hence the more consistent is its measurement of a trait. Consistency can be measured in four different ways and thus produce different types of reliability. The four types of reliability coefficients are:

1. Internal Consistency: correlating the scores achieved by individuals on the odd numbered items with their score on the even items.
2. Equivalence: correlating the scores the same individuals achieve on two forms of the same test.
3. Stability: correlating the scores the same individuals achieve from two administrations of the same test with the two administrations of the test separated by predetermined period of time.
4. Stability and Equivalence: correlating scores of the same individuals obtained on two forms of a test with the two administrations separated by a predetermined period of time.

9.2 Validity

Definition: Validity is the extent to which the test measures what it claims to measure. Validity also asks the question--What does this test measure?

EXAMPLE: An estimation of validity can be obtained by correlating test results with a criterion, as in correlating scores on the GATB with supervisor's rating of performance on the job.

It is important to remember that a test may be valid for one purpose but invalid for another and that a test may be reliable but invalid, that is, it can consistently measure something other than what it claims to measure. For example, a test designed to measure knowledge of statistics, may have items that students consistently miss, not due to lack of understanding of statistics but rather a lack of understanding of the vocabulary employed in the item. As with reliability there are four types of validity coefficients.

1. Content Validity: the extent to which subject matter experts judge that the test measures what is intended or the relevancy of items to objectives.
2. Concurrent Validity: the degree of correlation between test and criterion, for example, tests correlated with other tests or rating by an expert.
3. Predictive Validity: the ability of the test to predict an outcome. For example, an entrance examination ability to predict academic success.
4. Construct Validity: the extent to which a test adequately measures a theory or construct on which the test is based.

10.0 STANDARD ERROR OF MEASUREMENT

10.1 Standard Error of Measurement

Definition: The standard error of measurement (SE_m) is the standard deviation of the amount of error involved in measurement.

It should be apparent that no test is perfectly accurate. The test score is determined by the knowledge or ability of the individual taking the test and by the accuracy of the test in assessing the person's knowledge or ability.

Though there is no precise method of determining the accuracy of a given test, the standard error of measurement provides an estimate of the accuracy of a test.

Formula: $SE_m = SD\sqrt{1 - r_{xx}}$

Where SD is the standard deviation of a set of test scores, r_{xx} is the reliability coefficient.

Errors of measurement are assumed to be normally distributed, thus, an example of a test having a mean of 100, a SD of 15 and a reliability coefficient of .75 would yield a SE_m (using the above formula) or 8.0. In such a case all individuals whose true score is 100, approximately 68% will have test scores between 92 and 108, and 99% will have test scores between 76 and 124.

11.0 SUMMARY

The following is a condensed overview of the topics covered.

11.1 General Statistics

There are two major branches of statistics: descriptive and inferential. The purpose of descriptive statistics is to summarize data in order to show the regularities, trends, tendencies, and major relationships usually hidden in unordered raw data.

The purpose of inferential statistics is to draw conclusions about data that you don't have from data that you do have.

In both cases the goal is to extract all the meaning possible from the data you have.

11.2 Data

The input to statistical procedures is data. If the data are incorrectly identified and classified, or if they misrepresent the objects and events used for analysis, any numerical result from the statistical operation must be nonsense or worse, misleading. Data should be pronounced with a hard a, as in dāte-ah. Datum is the singular form of data; datum is, data are.

12.0 GLOSSARY

average--same as measure of central tendency. In more popular usage, same as mean or arithmetic mean.

average deviation (AD)--a measure of variability obtained by finding the arithmetic mean of the absolute values of the deviations of the individual scores from a measure of the central tendency of a distribution. Also called mean deviation (MD).

confidence interval--in a frequency distribution, the range on a variable within which a hypothetical score or statistic can be expected to fall with some specific relative frequency or probability--or confidence.

continuous-variable statistics--branch of statistics dealing with observations on variables that have continuous gradations of magnitude such as height, length, voltage, duration.

correlation--the relation between two variables; the extent of similarity in direction and degree of variations in corresponding pairs of scores on two variables.

correlation coefficient--(1) same as Pearson product-moment correlation coefficient; (2) any measure of relation between two variables. Also called coefficient of correlation.

datum, data--the record of an observation of an object or event in some behavioral context. In testing procedures, datum is synonymous with score.

decile--a name given to every tenth percentile. The 10th percentile is the first decile, the 20th percentile the second decile, and so on.

derived score--a score obtained from another score or scores by statistical methods; examples are a percentile score and a standard score. Also called transmuted or transformed score.

descriptive statistics--branch of statistics that summarizes data to show regularities, trends, and major relationships usually hidden in raw data. Treatment is limited to actual data available.

error of measurement--the difference between an obtained score and the corresponding true score.

frequency--the number of cases falling at any score, within any class of scores, or in each cell of a double-entry table.

frequency distribution--a tabulation of the frequencies of different scores or groups of scores when they are arranged in order of magnitude.

mean ($\bar{X}$ or M)--the sum of the scores divided by the number of scores; the point or value in a distribution of scores such that the sum of the deviations of the scores above it is equal to the sum of the deviations of the scores below it. Also called arithmetic mean and, in more popular usage, average or arithmetic average.

measure of central tendency--a value near the center of a frequency distribution that in some way typifies or represents the entire group of scores. Measures of central tendency include the mean, the median, and the mode. Same as average.

measure of dispersion--same as measure of variability.

measure of variability--a statistical measure that shows the extent to which the individual scores are concentrated about or spread out from the mean or other measure of central tendency. Also called measure of dispersion.

median (Md. or Mdn.)--the point or score value that divides the cases in a frequency distribution evenly into an upper and a lower 50 percent.

mode (Mo.)--the score with the highest frequency. Used as a quick but less precise approximation of central tendency than the mean or median.

N, n--the number of cases in a sample.

normal curve--same as normal probability curve.

parameter--some quantitative value that applies to a population or universe of interest. For example, the mean of a population is a parameter, the mean of a sample is a statistic.

Pearson product-moment correlation coefficient (r)--a number between the values of -1 and +1 that expresses the degree of relationship between two variables; the average product of standard scores on two variables; an index of the closeness of the grouping of cases about either regression line. Also called product-moment correlation coefficient, coefficient of correlation and Pearsonian correlation coefficient.

percentage--a fraction converted to a decimal with the decimal point moved two places to the right; the numerator value of a fraction with a denominator of 100.

percentile (P)--the point in a frequency distribution below which occur the percentage of the cases indicated by the particular percentile. Thus, 62 percent of the cases fall below the 62nd percentile.

percentile score--the percentile that corresponds to a given score in a frequency distribution; a given score expressed as a percentile.

probability--the relative frequency with which each alternative in a total set of alternatives appears defines the respective probability of the alternatives. The proportion of times a condition occurs relative to the total opportunities it has for occurring is its probability of occurrence. Probabilities may be established on empirical or theoretical grounds.

quartile--every 25th percentile. The first quartile is the 25th percentile; second quartile, the 50th percentile; third quartile, the 75th percentile. Sometimes, incorrectly, percentiles 1-25 are called the first quartile, 26-50 the second, 51-75 the third, and 76-100 the fourth.

range--the difference between the highest and the lowest obtained test scores for a group of subjects plus one unit: a crude measure of variability.

reliability--the accuracy with which a measuring device measures what it measures; verifiability of test results after a lapse of time or regardless of the particular items; consistency between results or repeated administrations of the same measuring device to the same individuals. Reliability is usually estimated in terms of the coefficient of reliability or the standard error or measurement.

reliability coefficient--an estimate of the correlation coefficient between a test and itself--that is, between the scores on a test given twice or on two equivalent tests given to the same subjects, without disturbance by such factors as memory, practice, boredom.

representative sample--a sample that duplicates the characteristics of the universe or population in all respects that are likely to influence results based on the sample.

sample--a number of cases selected from all the cases in a particular population or universe; part of a total population chosen for investigation. An attempt is often made to choose a sample that will be representative of the population of which it is a part. As a special case, the sample may consist of the entire population.

scaled score--a score derived from raw scores, which is designed to reflect certain characteristics not apparent in the raw scores. Examples are stanines, standard scores, and grade equivalents.

sigma--theoretically, a measure of the dispersion of scores about the mean of a population of scores. Its counterpart, the standard deviation, is the same measure when dealing with a sample of scores from the larger population. In practice the two terms are often used interchangeably. Usually, five or six sigmas (or standard deviations) cover the range of scores.

skewness (Sk.)--the extent to which the shape of a frequency distribution or curve departs from symmetry. When skewness is positive, the mean is ordinarily larger than the median and the mode; when negative, the mean is ordinarily smaller.

standard deviation (SD)--a measure of variability obtained by determining the square root of the mean of the squares of the deviations of the scores from their mean.

standard error--the estimated standard deviation of the values of a score or statistical measure that would be obtained if the measurement were repeated over and over again. The term is used when the deviations are thought of as errors, and the term standard deviation is used when they are not.

standard error of measurement--the estimated standard deviation of the differences between a series of obtained scores and their corresponding true scores. It can be estimated if the standard deviation and the reliability coefficient of the test are known.

standard score (z)--a deviation from the mean expressed in terms of the standard deviation of the distribution; a gross score minus the mean, divided by the standard deviation. Also called z-score and sigma score.

stanine--any one of the nine standard units formed when the distribution of scores is transformed into a normal distribution and then divided by 9, with the mean at 5 and a standard deviation of approximately 2. The extreme stanines (1 and 9) each include four percent of the scores; the amounts increase as the middle stanine is reached.

statistic--a computed value such as the arithmetic mean that applies to a particular sample.

summation (Σ)--addition. Summation is indicated by the sumbol Σ immediately to the left of the series being summed: ΣX means the sum of X values and ΣY^2 means the sum of the squares of Y values.

universe--all the cases in the class of things being investigated by statistical methods. The limits of a universe are specified by the characteristics common to all its members. For instance, the universe might contain white males alive in 1947 or white English males over 30 years of age and alive in 1947. Also called population.

validity--the extent to which a test or other variable measures what it is supposed to measure; the extent to which test scores can be used to predict criterion scores. Validity is often measured by the correlation of the variable with a criterion. Also called diagnostic value and discriminative power.

validity coefficient (r_{xy})--a correlation coefficient between a test (or tests) and a criterion that is intended to predict.

variability--the spreading out of the scores of a distribution from the mean or some other measure of central tendency.

variable--anything that can have different numerical values in different individual cases. Examples are height, temperature, test scores, percapita income.

variance--a measure of variability equal to the square of the standard deviation.

x--symbol for deviation from the mean, usually for a test or independent variable.

X--symbol for gross score, usually used for a test or independent variable.

z--symbol for standard score.

z-score--same as standard score.

13.0 PRE-TESTS

Section 1 -- GENERAL STATISTICS (Pre-test)

1. Whenever we use a number (for example, an average) to describe a distribution, we are using

   _____ statistics.

2. The mean, the median, the mode, and the standard deviation are all

   _____ _____.

   They simply

   _____ the distribution of scores.

3. All of the following are aspects of descriptive statistics except:

   a. variability

   b. inference

   c. correlation

   d. central value

4. The accuracy of statistical inferences depends upon:

   a. the critical ratio

   b. the size and representatives of the sample

   c. the standard error

   d. the standard deviation

Answers:

1. descriptive
2. descriptive statistics, describe
3. b
4. b

Section 2 -- CENTRAL TENDENCY (Pre-test)

1. The arithmetic average is the:

   a. mean

   b. median

   c. mode

   d. standard deviation

2. The middlemost score in a distribution is the:

   a. mode

   b. median

   c. mean

   d. arithmetic average

3. The mode of a distribution is the:

   a. arithmetic average

   b. middlemost score

   c. score with the greatest frequency

   d. standard deviation

4. The various properties of the normal curve can be used for statistical purposes:

   a. if the standard deviation is large

   b. if the standard deviation is small

   c. when the distribution approximates a bell-shaped curve

   d. regardless of the nature of the distribution

5. "Chance" events yield a (n):

   a. skewed distribution

   b. bimodal distribution

   c. normal distribution

   d. inverted J-curve

6. John has taken eight tests during a semester, each of which counts the same toward his final grade. To get an idea of how well he has done, he adds up each numerical grade and divides by eight. He has computed a

   _____ grade.

7. If all of John's scores are arranged from low to high and count up (or down) until we come to the middle score, we have obtained the

   _____ score.

8. Whenever we are talking about the score that is received most often, we are talking about the

   _____.

9. Compute the mean, median and mode for the following distribution of scores: 5, 4, 4, 5, 2, 9, 9, 8, 2, 9, 25, 46, 26.

   mean = _____

   median = _____

   mode = _____

10. The point below which 50 percent of the scores fall is the:

    a. mean

    b. median

    c. mode

    d. 49th percentile

Answers:

1. a
2. b
3. c
4. c
5. c
6. mean
7. median
8. mode
9. mean = 12 5/6; median = 8; mode = 9
10. b

Section 3 -- RANDOM, CONTINUOUS AND DISCRETE VARIABLES

1. A random variable is one whose

   _____ value is not controlled by the experimenter when drawing samples from the population.

2. Observing heads or tails of a tossed coin is an example of a random variable.

   True or False

3. In an attempt to assess public opinion, sampling every other registered Republican voter is an example of a random variable.

   True or False

4. A continuous variable is one with a

   _____ gradation of magnitude.

5. The bathroom scale is an example of a continuous variable.

   True or False

6. A thermometer is an example of a continuous variable.

   True or False

7. A discrete variable is one in which changes occur in

   _____ or _____ steps.

8. A conventional light switch is an example of a discrete variable.

   True or False

Section 3 -- RANDOM, CONTINUOUS AND DISCRETE VARIABLES/Pre-test (Continued)

9. A true-false question is an example of a discrete variable.

   True or False

10. Marital status is an example of a discrete variable.

   True or False

Answers:

1. numerical
2. T
3. F
4. continuous
5. T
6. T
7. all, none
8. T
9. F
10. In this case, since there are many stages and conditions of marital status, there is no correct T or F response.

Section 4 -- VARIABILITY (Pre-test)

1. Statistics is a form of mathematics which has been developed to analyze and evaluate:

    a. measurements

    b. differences

    c. continuous variables

    d. discrete variables

2. Another name for variability among a set of scores is:

    a. dispersion

    b. scatter

    c. a and b

    d. none of these

3. Variability in scores can result from:

    a. giving a test to each member of a group

    b. giving the same test at several different times to the same individual

    c. using a test which is subject to mistakes in scoring

    d. all of these

4. Scores on a published intelligence test for an individual are more stable over time than scores on published personality tests. The principal reason for this is that:

    a. intelligence tests are better designed

    b. intelligence tests are easier to score

    c. other personality traits are more complex than intelligence

    d. intelligence is a more stable trait than other personality traits

5. A graph or plot which indicates how many individuals received a particular score is called a:

   a. frequency distribution

   b. variability display

   c. scatter plot diagram

   d. histogram

6. A symmetrical bell-shaped curve describes the shaped curve describes the shape of a:

   a. bimodal distribution

   b. normal frequency curve

   c. skewed distribution

   d. set of time scores

7. A flat distribution curve suggests:

   a. relatively wide individual differences

   b. relatively narrow individual differences

   c. a narrow range of scores

   d. an inaccurate measuring instrument or test

8. If one gave a mechanical aptitude test to a group of high school-girls and to a group of mechanical engineers, and then combined the scores, the resulting frequency distribution would be:

   a. bimodal

   b. rectangular

   c. normal

   d. bell-shaped

9. If one gave a very easy multiple-choice test to a group of individuals, the resulting frequency distribution of scores would probably be:

   a. bimodal

   b. symmetrical

   c. skewed

   d. rectangular

Answers:

1. b
2. c
3. d
4. d
5. a
6. b
7. a
8. a
9. c

Section 5 -- MEASURING VARIABILITY (Pre-test)

Computation of Standard Deviation and Standard Score

The most important and frequently used measure of variation, the standard deviation, is calculated by summing the squares of the deviations about the mean, dividing by N, and taking the square root. Furnish your own values for X.

$$S_x = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N}}$$

Subject	X	$X - \bar{X}$	$(X - \bar{X})^2$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

$\Sigma X =$ $\Sigma(X - \bar{X})^2 =$

$\bar{X} =$

$\bar{X} = \frac{\Sigma X}{N}$

After having learned to compute the standard deviation, computation of the standard score involves only one additional step. The standard score is derived by taking deviations from the mean and dividing by the standard deviation.

$$z_x = \frac{X - \bar{X}}{S_x}$$

Section 6 -- SAMPLING (Pre-test)

1. When we carry on research we usually observe, test or survey only some members of a larger group of subjects or cases. This large group is referred to as the:

   a. sample

   b. distribution

   c. population

   d. criterion group

2. Unbiased samples drawn from a population usually differ in the characteristic being measured. The explanation for these differences is that they are due to:

   a. random sampling

   b. unknown factors, i.e., chance

   c. careless sampling

   d. inferences about the population

3. Because unbiased samples carefully drawn from a population differ unpredictably, one should usually try to measure all the individuals in a population.

   True or False

4. The more widely the scores on a given characteristic vary in a population the ____ the sample size should be.

   a. larger

   b. smaller

   c. sample size is irrelevant if sampling is unbiased

5. Not all samples are drawn randomly. Some are selected by systematic sampling such as:

   a. choosing every 25th name in the telephone book

   b. interviewing the head of the household living on the northeast corner of the block in a residential area

   c. testing each person whose best two digits of his ID number end in 15

   d. all of these

6. In defining a particular population, it is essential to describe precisely:

   a. the common and unique characteristics of those individuals who make up the population

   b. the methods of measurement

   c. the conditions under which measurements are taken

   d. all of these

7. When we wish to estimate one or more attributes or characteristics of a population from some of its members, we resort to:

   a. sampling

   b. descriptive statistics

   c. factor analysis

   d. hypothesis testing

8. In biased sampling there is:

   a. a systematic error

   b. a condition which will always bias the results of your investigation

   c. a stratification of sampling

   d. all of these

9. When we wish to draw a random sample, our main concern is to make sure that chance factors are:

   a. eliminated

   b. allowed free play

   c. brought under precise control

   d. exposed

Answers:

1. c
2. b
3. F
4. a
5. d
6. d
7. a
8. a
9. b

Section 7 -- DERIVED SCORES (Pre-test)

1. When we transform a set of scores from one form to another, e.g., from raw score to percent correct, we have in the new form, a set of

   _____ scores.

   a. percentile

   b. standard

   c. derived

   d. simplified

2. All percentiles are percentages but not all percentages are percentiles.

   True or False

3. If we have 35 subjects and 28 passed a screening test and seven failed, we cay say that:

   a. 80% passed

   b. 20 percent failed

   c. 100% took the test

   d. all of these

4. The percentage of individuals in a group of individuals that equal or exceed a given score is called:

   a. the percentage correct

   b. a percentile

   c. the success rate

   d. the criterion

5. For a given distribution of scores the 50th percentile has always the same value as the:

   a. mean

   b. median

   c. mode

   d. standard deviation

6. The eighth decile is equal to the 80th percentile.

   True or False

7. On the chart showing the normal curve, the derived score designated as the 16th percentile is equivalent to a T-score of:

   a. 7

   b. 85

   c. 40

   d. 12%

8. Which of the following scores are not based on the standard deviation?

   a. stanines

   b. College Entrance Examination scores

   c. z-scores

   d. percentiles

9. In order to compute standard scores, we need to know the _____ of a distribution.

   a. mean

   b. standard deviation

   c. sample size

   d. a and b

10. T-scores are standard scores with a mean of 50 and a standard deviation of 10. To convert raw scores to T-scores, one first must compute:

    a. Z-scores

    b. percentile equivalents

    c. stanines

    d. the product-moment correlation

Answers:

1. c
2. a
3. d
4. b
5. b
6. T
7. c
8. d
9. d
10. a

Section 8 -- COVARIATION (Pre-test)

1. To compute a correlation coefficient, one must have at a minimum:

   a. two sets of scores on the same individuals

   b. comparable scores on two different groups of individuals

   c. an experimental group and a control group

   d. more scores on the x-axis than on the y-axis

2. One would expect the correlation between students' grades and their number of absences from class to be:

   a. positive

   b. negative

   c. random

   d. none of the above

3. A correlation coefficient expresses:

   a. the variability of two variables

   b. the relationship between two variables

   c. the standard error of two variables

   d. the difference between two variables

4. A scatter diagram with points running from the bottom left to the top right along a diagonal line will yield:

   a. a negative r

   b. a negative rho

   c. an r close to 1.00

   d. a rho close to 0.00

5. Of the following correlation coefficients, the one indicating the greatest relationship is:

   a. .90

   b. .50

   c. -.95

   d. -.50

Answers:

1. a
2. b
3. b
4. c
5. c

Section 9 -- RELIABILITY-VALIDITY (Pre-test)

1. Validity has usually meant:

   a. the dependability of a test

   b. confirming the null hypothesis

   c. how well a test measures what it is supposed to measure

   d. testing the significance of a difference

2. Reliability can be determined by:

   a. checking against some outside criterion

   b. comparing ranges of two halves of a test

   c. correlating the scores of two distinct types of test

   d. correlating two halves of the same test

3. When a test gives the same scores on repeated administration to the same individuals, it has high:

   a. reliability

   b. validity

   c. causality

   d. none of these

4. How well individuals do on a job may be used as a criterion for examining a test's:

   a. significant difference

   b. validity

   c. reliability

   d. bias

5. The comparison of scores made by members of a group whose abilities in the area tested are known, with the scores of a random sample of the population, is an effective method of test validation.

   True or False

Answers:

1. c
2. d
3. a
4. b
5. T

Section 10 -- STANDARD ERROR OF MEASUREMENT (Pre-test)

1. The standard error of measurement can be used to indicate the precision or accuracy of a test. The more precise or accurate the test, the

   _____ will be the standard error of measurement.

   a. larger

   b. smaller

   c. more variable

2. The standard error of measurement is more closely related to the

   _____ of a test.

   a. reliability

   b. validity

   c. consistency in results

   d. a and c

Answers:

1. a
2. d